
Data Management for XENON1T and South Pole Telescope

DOMA in Astronomy, Genomics and High Energy Physics

Nov 16, 2017

Benedikt Riedel
Lincoln Bryant
Rob Gardner
Judith Stephen
University of Chicago



Open Science Grid



XENON1T



- 1st ton-scale experiment
- 3.2t of LXe, 2t in TPC
- All systems commissioned since Fall 2016
- Calibration and science data taking now ongoing



XENON collaboration



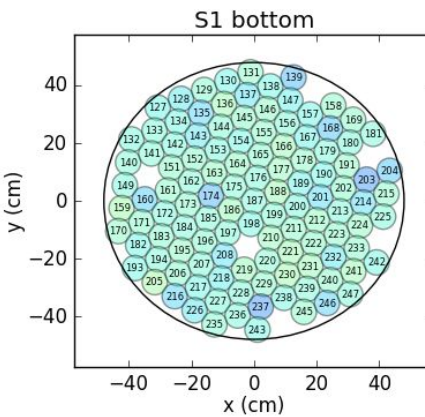
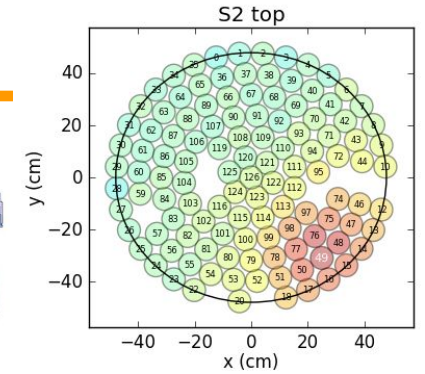
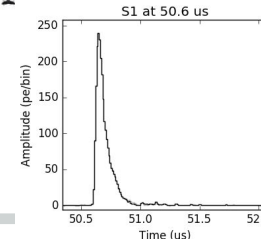
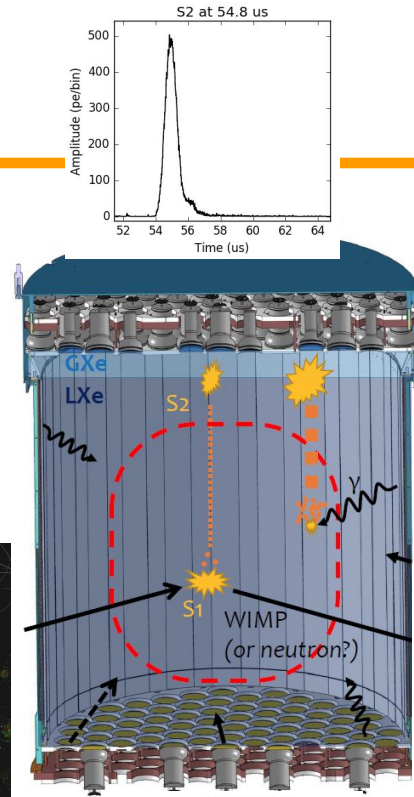
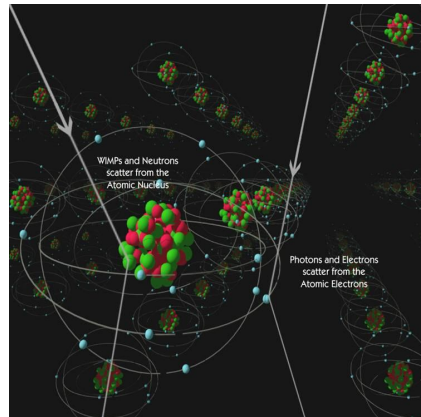
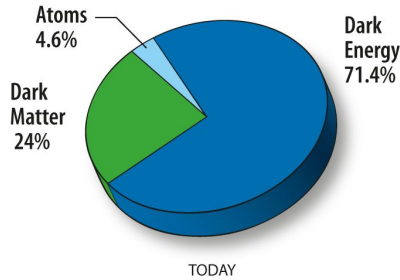
Logos of participating institutions:

- Columbia
- RPI
- Nikhef
- Stockholm University
- Muenster
- Mainz
- THE UNIVERSITY OF CHICAGO
- UCLA
- UC San Diego
- UCSD
- Rice
- Purdue
- Coimbra
- Subatech
- LPNHE
- Bologna
- LNGS
- Torino
- MPIK
- UNIFREIBURG
- University of Zurich
- Zurich
- جامعة نيويورك أبوظبي
- NYU | ABU DHABI
- NYUAD
- פנח ויצמן למדע
- WEIZMANN INSTITUTE OF SCIENCE
- Weizmann

XENON1T Science

- No idea what 95.4% of universe is made of
- Looking for Dark Matter candidate - Weakly Interactive Massive Particles (WIMP) - through nuclear recoil on liquid Xenon
- WIMP nuclear recoil produces charge and time signature in the detector readout that can easily distinguish nuclear recoil from other interactions
- First results:

<https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.119.181301>



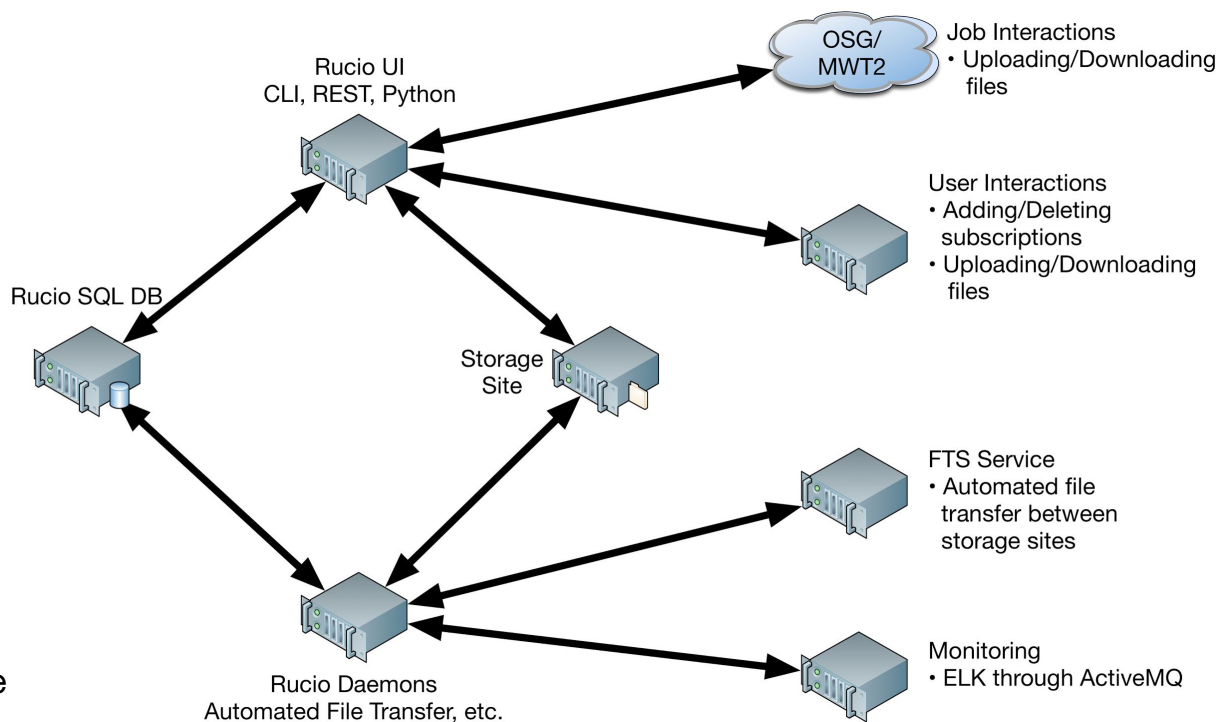
Xenon1T Storage and Processing Challenge

- Storage allocated at European Grid Infrastructure (EGI) and Open Science Grid (OSG) sites - Not enough storage at any one site for all the data
- Computing and storage on OSG and EGI sites through single interface for each
- Could not use Globus Online to automate transfer to/from EGI sites
- How to manage the data?

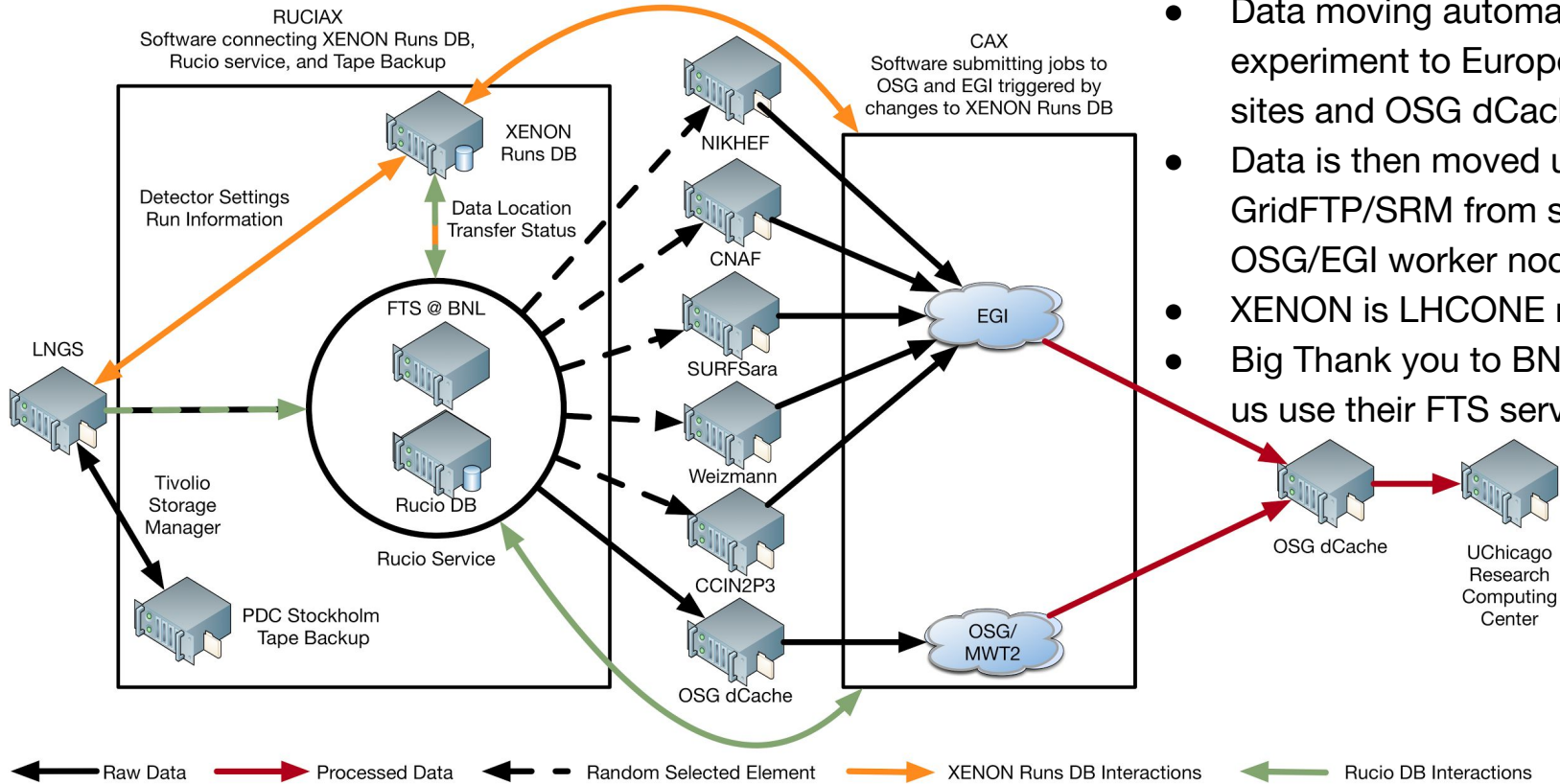


Rucio

- Data Management software created by the ATLAS experiment at the LHC, used by XENON1T, AMS, and ATLAS
- Provides single interface to distributed data
- Automated replication of data through a “subscription” model, i.e. a site is “subscribed” to a certain data set
- Built with future in mind, i.e. scalable database infrastructure, “common” data transfer methods support (GridFTP, SRM, XrootD, S3, etc.), monitoring through ELK, etc.
- Somewhat ATLAS-centric features present; Development to accommodate other experiments on-going



Xenon1T Data Management



- Data moving automatically from experiment to European storage sites and OSG dCache
- Data is then moved using GridFTP/SRM from storage to OSG/EGI worker nodes
- XENON is LHCONE member
- Big Thank you to BNL for letting us use their FTS service

XENON1T Statistics

Rucio Statistics:

- > 1.2M Files
- ~16k Datasets
- 9 storage endpoints
- 1887.5 TB of available storage
- 854.1TB of available storage used
- Adding 1.3 TB per day, 200+ files per hour
- > 115 GB per hour transferred

Future Plans

“Rucio as a Service”

- Rucio can run from Docker images
- Spawn up Rucio instances as needed for experiments and researchers
- Setup single central DB instance with sub-DB for each Rucio instance
- Creating OSG FTS instance

SPT Experiment

- Microwave-Millimeter telescope located at the South Pole
- Upgrades in 2016/2017:
 - SPT 3G - Third generation detectors installed
 - 10x detectors = 10x data rate
 - Added detection channel
 - New compute and storage hardware at South Pole and UChicago



The SPT-3G Collaboration (Feb. 2016)

~70 scientists, across ~20+ institutions

Funded By:



Science

- Cosmic Microwave Background: CMB power spectrum, CMB Lensing, CMB Polarization, Epoch of Reionization
- Galaxy Clusters: Cluster Cosmology, Cluster Catalog
- Galaxy Evolution: High- z strongly-lensed star-forming galaxies, Catalog of mm sources

Data Management Challenges

From the instrument to archival:

- Limited bandwidth/high latency between experiment and researchers
- Online data acquisition and on-site processing
- Transport buffering & archival or raw (precious) data
- On-going processing and analysis
- Accessing archival storage

From central storage to OSG jobs:

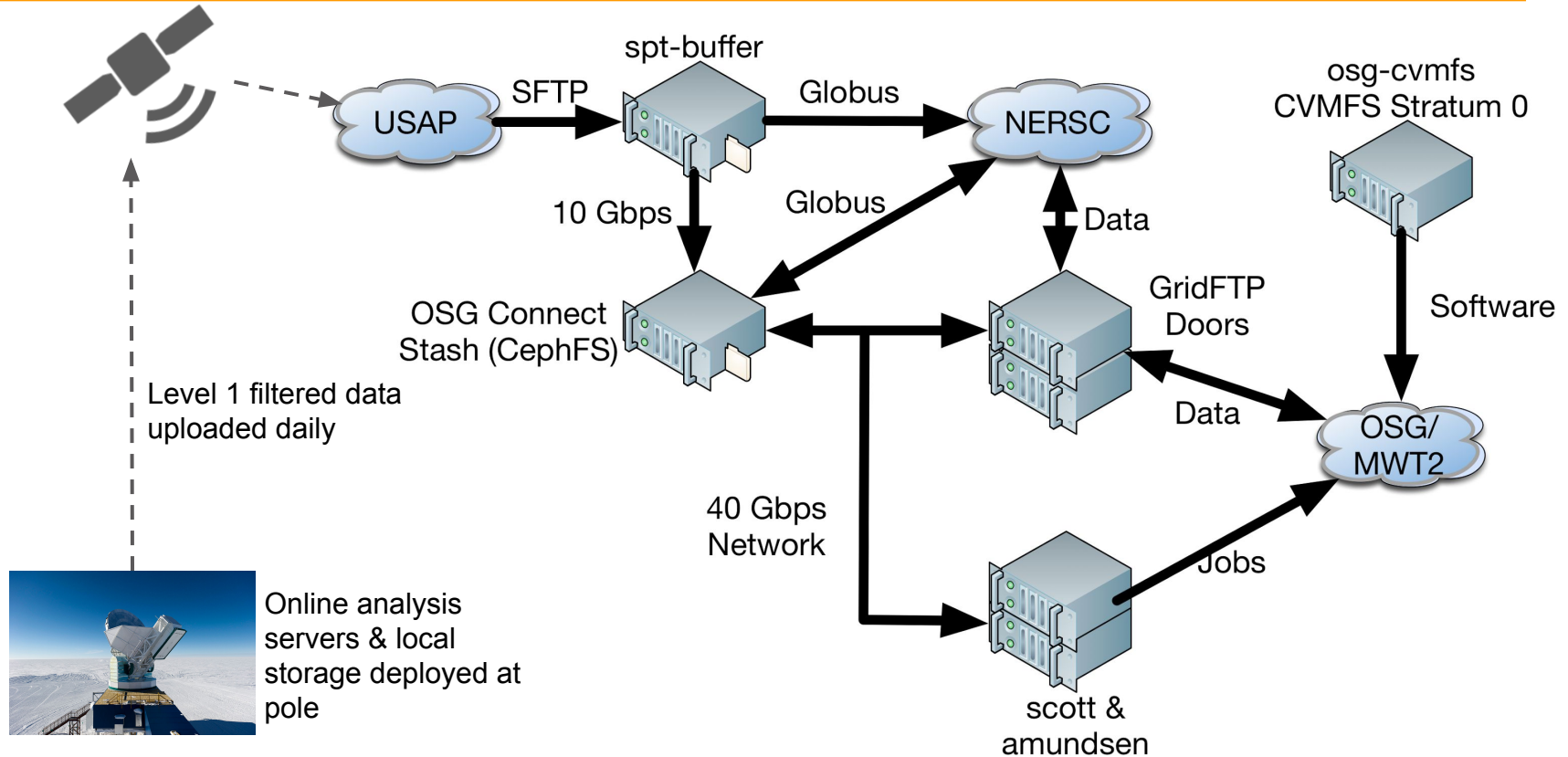
- Single central store (for now)
- Filesystem limits
- Limits on GridFTP bandwidth

Nearline Data Infrastructure - South Pole



- **New Hardware in red** (no Fry's down the road edition):
 - 4x Dell R730s:
 - 2x R730 for analysis work (HTCondor pool)
 - 1x R730 as hypervisor
 - 1x R730 hot spare
 - 2x Dell R330s: Storage controller + backup
 - 2x Dell MD1280s:
 - Primary Copy: ZFS pool, 42x 8 TB, NFS mounted to all R730s
 - Secondary Copy: JBOD, 28x 8 TB
 - 2x UPSes, 6x PDUs
- Old hardware in green - Part of online analysis HTCondor pool

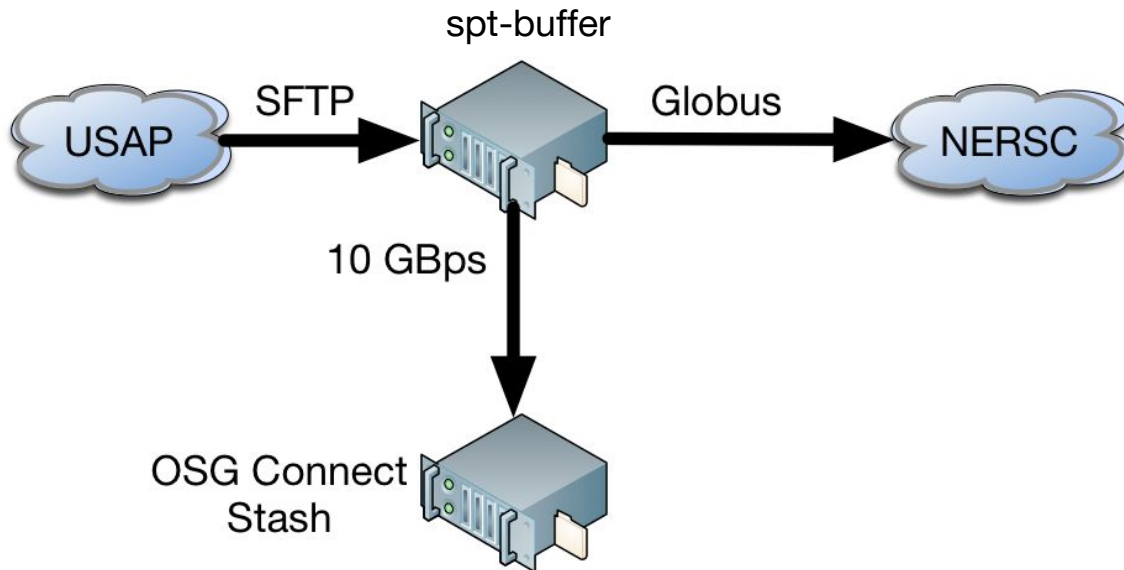
Transport & Processing: Pole to the US



Data Infrastructure: Pole to OSG & NERSC

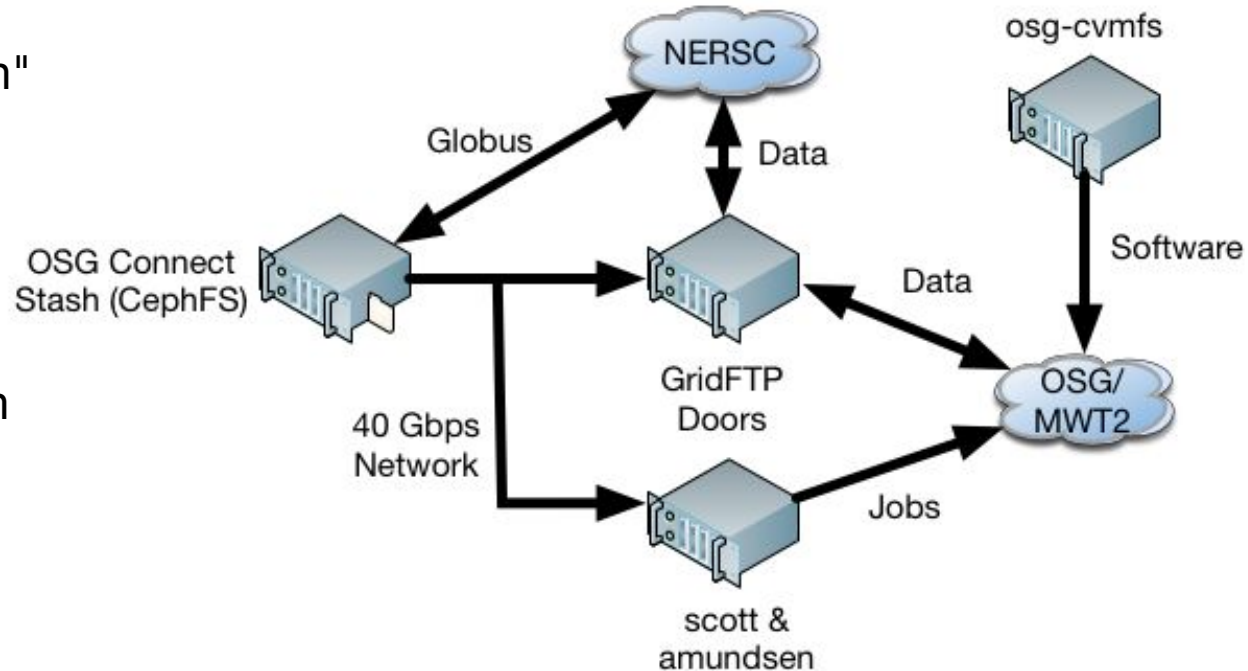
- Data Ingest and Archival Infrastructure:

- spt-buffer:
 - VM with 4 TB disk attached
 - Total 40 day buffer
- /spt
 - Part of OSG Stash (CephFS)
 - rsync'ed from spt-buffer
- Backup from spt-buffer to NERSC HPSS, partially automated



Data Infrastructure for Software & Analysis

- Analysis Servers
 - "scott" & "amundsen"
 - Dell R630s
 - Login, interactive, Condor, JupyterHub
 - 10Gbps
 - /spt from OSG Stash
- 2x gridftp with bonded 2x10 Gbps
- osg-cvmfs - SPT OASIS Stratum-0



Future Changes

Moving away from CephFS to Ceph Object Store for raw data

- “Closer to the metal” - No bottleneck from metadata server
- Built-in scalable file transfer protocol
- No more tiered storage - Not necessarily required if running newest release
- A lot of momentum behind S3/Swift-like file transfer
- POSIX-like access through NFS